Universal Prevention of Child Behavioral Disorders by the TRIPLE P-Parent Training. 10-Year Effectiveness from Mothers', Fathers', and Adolescents' Perspectives

Kurt Hahlweg & Wolfgang Schulz
Technische Universität Braunschweig
Institut für Psychologie, Abteilung für Klinische Psychologie, Psychotherapie und Diagnostik, Humboldtstr. 33, D-38106 Braunschweig

Author Note

Please address correspondence to Prof. Dr. Kurt Hahlweg, Schwäbische Str. 7, 10781 Berlin, Germany. E-Mail: k.hahlweg@tu-bs.de

This study was carried out within the framework of the DFG project "*Zukunft Familie III*" ["Future Family III"] (HA 1400/17-1, 2: *Längsschnittstudie zur Vorhersage von Verhaltensauffälligkeiten im Jugendalter unter Berücksichtigung von Risiko- und Schutzfaktoren* [Longitudinal study for the prediction of behavioral disorders during adolescence, Taking risk and protection factors into consideration]).
We thank the families for their participation and Tim Bothe (BSc (Psychology) for his helpful statistical analysis.

Conflicts of Interest: Kurt Hahlweg is member of the Triple P International Scientific Advisory Committee. Kurt Hahlweg and Wolfgang Schulz do not have any financial interests.

Please cite: Hahlweg, K. & Schulz, W. (2018). Universelle Prävention nach 10 Jahren aus Sicht der Mütter, Väter und Jugendlichen. [Universal Prevention of Child Behavioral Disorders by the TRIPLE P-Parent Training. 10-Year Effectiveness from Mothers', Fathers', and Adolescents' Perspectives]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 47*, 1-15.
https://doi.org/10.24355/dbbs.084-201901071038-0

Abstract

Background: Mental problems in children are widespread and cannot be reduced through treatment only. Prevention is therefore urgently needed although it is unclear how effective such strategies may be, particularly in the long term. **Aim***:* Can a parent-centered universal prevention program that is effective in the short term also yield effects after 10 years? **Method**: According to their preschool location, N = 477 families were randomly assigned to the parent training prevention program (Triple P Positive Parenting Program, TP; Sanders, 2012) or the control group (CG). In all, 77% accepted the TP offer (T+), while 23% declined it (T-). The 10-year effectiveness of the program was established with self-report measures of mothers, fathers, and adolescents from N = 361 families. Results: The intention-to treat analysis (comparison TP vs. CG) yielded negligible findings. By contrast, the differential analysis from the T+ mothers' perspective found long-term improvements in Child Behavior Checklist (CBCL) internalizing and externalizing behavior and relationship satisfaction in comparison with CG and T- mothers. At 10 years, compared with pre-assessment, T+ mothers reported the smallest increase in the CBCL sum score of internalizing and externalizing behaviors, 5%, while CG (20%) and T- (33%) mothers reported far higher rates. Contrary to the hypotheses, parenting behavior did not change over time. T+ fathers reported improvements in parenting behavior, while adolescents reported negligible outcomes. **Conclusion**: The results support the long-term effectiveness of the TP program as a universal prevention intervention, at least from the T+ mothers' perspective. More research should be conducted with the T- families because they showed worse outcomes than the control group.

Key words: Universal prevention, parent training, Triple P, behavior problems, adolescents, longitudinal study

Universal Prevention of Child Behavioral Disorders by the Triple P-Parent Training. 10-Year Effectiveness from Mothers', Fathers', and Adolescents' Perspectives

In western industrial societies up to 20% of children and adolescents have mental disorders and behavioral problems (Belfer, 2008; Kieling et al., 2011), which burden those affected, their families and the social environment and thus have a high social relevance. In Germany there are similar prevalence rates as revealed by the representative KIGGS trials of the child and adolescent health survey of the Robert Koch Institute (Ravens-Sieberer, Wille, Bettge & Erhart, 2007; Hölling, Schlack, Petermann, Ravens-Sieberer & Mauz, 2014). These findings indicate that a longer-term reduction of the prevalence rate calls for targeted preventive measures within a public health framework.

For parents, these disorders often represent a considerable burden and often lead to increased partnership conflicts and separations (e.g. in ADHD: Wymbs et al., 2008). For the affected children and adolescents, a mental illness with significant impairments in the areas of school, social relations and somatic health is associated with the risk of chronicity (O'Connell, Boat & Warner, 2009).

Economic consequences. These long-term negative consequences create significant societal costs (direct costs for psychotherapeutic and psychiatric interventions, indirect costs for the often diminished achievements in school and occupation). In addition, there is an increased likelihood of delinquency, alcohol and drug addiction and sexual risk behavior (Caminis, Henrich, Ruchkin, Schwab-Stone & Martin, 2007; Englund, Egeland, Oliva & Collins, 2008; Farrington & Welsh, 2007; Gustavsson et al., 2011; O'Connell et al., 2009; Pingault et al., 2013). Far too often, there is no adequate treatment of childhood mental disorders which is due to insufficient healthcare structures. For the reasons mentioned above, the prevention of long-term impairments and the establishment of appropriate prevention programs such as parenting or problem-solving courses for schoolchildren should be the goal of every healthcare system.

Risk factors. The development of mental health problems in childhood is multifactorial in the sense that both biological and psychosocial risk factors play an important role (O'Connell et al., 2009). In particular, those empirically secured psychological factors are crucial and need to be fundamentally changed for the prevention of childhood behavioral disorders by means of parenting courses. These

factors include inconsistent and punitive parenting, negative family communication patterns, partnership conflicts and mental disorders of the parents, such as depressive disorders or alcohol dependence. The reduction of at least family risk variables through preventive interventions as early as possible in early childhood is urgently required: The sooner intervention takes place, the greater the chance that the behavior has not already stabilized and has become chronic (Hahlweg & Heinrichs, 2007; Heckman, 2008).

Insecurity of parents. Another important reason for the increased use of parenting courses is that many parents are very insecure with regard to their parenting skills. So, for example, 68% of 850 mothers of nursery-school children reported that they did not know if they were performing their parenting tasks well or poorly (Braunschweiger Kindergartenstudie [Kindergarden trial]; Kuschel et al., 2004).

Parenting skills courses. Especially cognitive-behavioral parenting courses have proven to be effective (Furlong et al., 2012; Nowak & Heinrichs, 2008; Piquero et al., 2016; Sandler, Schoenfelder, Wolchik & MacKinnon, 2011; Stattin, Enebrink, Özdemir & Giannotta, 2015; Weiss, Schmucker & Lösel, 2015; Weisz & Kazdin, 2010) and are regarded as the gold standard in the field of prevention of behavioral problems in children (United Nations Office on Drugs and Crime, 2009; World Health Organization, 2009). In a meta-analysis, which comprised 77 trials about parent trainings with different theoretical orientations, Kaminski, Valle, Filene and Boyle (2008) reported a low effect size (d) of 0.34 (95% confidence interval [0.29; 0.39]). Intervention components that had a major effect, included an increase in positive parent-child interactions, emotional communication skills and the learning of time-out as a parenting strategy as well as consistent parenting skills and hands-on practice of new skills during parenting course sessions with one's own child (Kaminski et al., 2008).

Lundahl, Risser und Lovejoy (2006) included 63 randomized controlled trials and yielded low to moderate effect sizes for the reduction of child behavior problems (d = 0.42), parenting behavior (d = 0.47) and an increase of positive attitudes to parenting behavior (d = 0.53). According to O'Connell et al. (2009), evidence-based programs include the Positive Parenting Program - Triple P (Sanders, 2012), the Incredible Years Program (IY, Webster-Stratton, 1998), the Parent-Child Interaction Therapy (PCIT, Fernandez & Eyberg, 2009) and Parent Management Training –

Oregon Model (PMTO, Forgatch & Patterson, 2010). All of these programs are based upon social-cognitive learning models and connect them with developmental psychological theories (Sanders, Kirby, Tellegen & Day, 2014).

A key finding of all these secondary studies is that parenting skills training is effective in changing dysfunctional parenting behavior and is also often effective in reducing child behavioral problems. However, the long- term effectiveness of parent trainings as a universal preventive measure aimed at all parents and not just parents of children who already display behavioral problems (= indicated prevention), has rarely been studied, as most studies have follow-ups of only one year (Averdijk, Zirk-Sadowski, Ribeaud & Eisner, 2016; Hiscock et al., 2008). An exception is the "Erlangen-Nürnberger Development and Prevention Study" by Lösel, Stemmler and Bender (2013). This control group study comprised N = 609 families with a 3 to 5-year-old child. In the follow-up after 5 years, effect sizes between d = 0.10 and 0.30 were found, depending on the outcome variables examined.

Background of this study

The aim of the present study is to investigate the results of a parent-centered universal preventive measure (Triple P) *10 years* after its implementation and thus, this study is the first of its kind – also internationally. The basis for the catamnesis were two studies (ZF I and ZF II), which were carried out in our research group from 2000-2006. In the studies, the Triple P (Positive Parenting Program, Sanders, 2012, see below) was used. This parenting program exists in at least 30 countries around the world and has been examined in over 100 studies involving approximately 16.000 families.

A meta-analysis by Sanders et al. (2014) showed significant, short-term small to medium (pre-post) effect sizes for social, cognitive and emotional child behavior (d = 0.47), parenting (d = 0.58), parental satisfaction (d = 0.52), parental psychological burden (0.34), and partnership satisfaction (d = 0.23). However, there are also studies in which Triple P interventions did not achieve significant results, such as Averdijk et al. (2016) or Wilson et al. (2012).

On the one hand, Triple P represents a parenting program, on the other hand the term also stands for an entire system of family support. Triple P as a system comprises five levels of intervention on which parents can seek support (Heinrichs & Hahlweg, 2009; Sanders, 2012). The levels vary in their intensity of support, because one of the basic assumptions of Triple P is that not every family has the same need

for support and advice. In our studies, Triple P parenting courses were carried out in a group format (Group Triple P; Level 4; Sanders, 2012).

Parents were introduced to a comprehensive range of 17 parenting skills in four two-hour sessions in the following areas: Basics of positive parenting, possible causes of problem behavior, parenting strategies to promote positive child development, and dealing with problem behavior. After the four group sessions, parents had the opportunity to have three to four weekly individual telephone sessions (15-20 minutes each) to discuss progress, questions and difficulties with the Triple P provider. This contributes to the stabilization of implemented strategies and supports the generalization of future problems.

In this study, data from two longitudinal studies were evaluated. The randomized control group study called "**Zukunft Familie I**" (ZFI, funded by the German Research Foundation DFG, HA 1400 / 1–5), focused on an evaluation of the 4-year effectiveness of the parenting program Triple P for the universal prevention of behavioral problems of children aged 3 to 6 years. N = 280 families were recruited from 17 randomly selected day-care centers in Braunschweig. N = 94 families were assigned to the control group (CG), N = 186 were offered to participate in the Triple P program. N = 144 families (T+, 77%) accepted this offer, N = 42 families (T-, 23%) refused to participate in the parenting course. The families refusing the training did not differ from those accepting the program with regard to socioeconomic variables, but with regard to the problem behavior of children: Families who decided to take part in the course reported more behavioral problems than those who refused (Heinrichs, Hahlweg et al., 2006).As dependent variables a combination of interviews and standardized questionnaires was chosen. Data collection was done at five time points: pre, post (immediately after completion of training), follow-ups 1 to 4 (after 1, 2, 3, and 4 years) after the initial survey (for further details see Hahlweg, Heinrichs, Naumann, Kuschel & Bertram, 2010; Heinrichs, Bertram, Kuschel & Hahlweg, 2005; Heinrichs, Hahlweg, Bertram, Kuschel & Naumann, 2006; Heinrichs et al., 2009; 2017).

Of the TP+ mothers, 89% participated in at least 3 of the 4 sessions. Five accredited providers led a total of 28 groups. Parents regarded the Triple P courses as very helpful: Regarding the quality of the training, 90% of parents gave a rating of good to excellent (Heinrichs, Hahlweg et al., 2006).

In the uncontrolled study **Zukunft Familie II** (ZFII, funded by the Jacobs Foundation, Zurich; Heinrichs, Krüger & Guse, 2006), 197 families from socially disadvantaged urban areas in Braunschweig were offered the Triple P course. Data collection took place at four times of measurement (pre, post, FU 1 and 2 years, see Fig. 2). The aim of the study was, on the one hand, to examine whether families could benefit from different financial incentives, and on the other hand, whether different settings (Group vs. Standard Triple P which is delivered individually) have an impact on the effectiveness. All participants were offered a free Triple P course, either paid or unpaid and as a Group or Standard course. A total of 17 Group and 113 Standard courses were conducted by six accredited Triple P providers. 85% of the mothers participated in at least 3 of the 4 sessions. The parenting course was found to be very helpful by parents: Regarding the quality of the course, over 90% of the parents gave a rating of good to excellent (Heinrichs, Krüger et al., 2006). For further details on the sample and data collection see Heinrichs, Krüger et al. (2006).

In both studies inclusion criteria were child age (2.6 to 6.5 years) and a basic understanding of the German language (e.g. can understand German and can speak it at least a little bit). Proficiency in writing German was not required, as this could not be attested to all German families either.

Study Zukunft Familie III (ZFIII)

Objectives and aims. The primary objective of the ZFIII-study was to evaluate the long-term effectiveness of the Triple P parenting course on the basis of a universal sample over a follow-up of 10 years (FU5). The central question was: Will Triple P still have positive effects from the perspective of mothers, fathers and young people 10 years later? Compared to the control group and the T- group, we expected an improvement in parenting behavior and a long-term reduction of external and internal problems through the universal use of the parenting course (primary criteria). In addition, a long-term reduction of the individual burden of parents was expected as well as a long-term improvement of partnership satisfaction (secondary criteria). Since this is a study in the field of universal prevention, low to moderate effect sizes were expected.

Method

Recruitment. The total sample consisted of N = 477 families (study ZFI: n = 280; study ZFII: n = 197). For n = 221 Families (n = 63 ZFI, n = 158 ZFII) the addresses were wrong and had to be retrieved by contacting the residents'

registration offices. If unsuccessful, Internet research via Facebook and other sources were used. Lastly families with the same last name were contacted by telephone. After identification, parents were contacted by telephone; if they agreed to participate, the following questionnaires were mailed to them: Child Behavior Checklist 4–8 (CBCL; Döpfner et al., 2014), Depression-Anxiety-Stress Scale (DASS; Köppe, 2001) and the Parent Scale (EFB; Naumann et al., 2010).

The data collection was done with a combination of interviews and a written and electronically standardized survey which lasted approximately 2.5 hours. Interviews were carried out within the framework of home visits or on the premises of the Technical University of Braunschweig. The parent in question (in 94 % of the cases, the mothers) and the adolescents were interviewed in parallel sessions, separate from each other. The interviews were conducted by two interviewers (at least one of them with a Master's degree in Psychology in training to become a clinical psychologist [German: Psychological psychotherapist]). Since some of the items were sensitive, the adolescents were, as a rule, questioned by an interviewer of the same gender. All interviewers were trained intensively were licensed to conduct diagnostic interviews for emotional disorders in childhood and adolescence (Kinder-DIPS) (Schneider, Unnewehr, & Margraf, 2009). The families received a compensation of a total of 80 € (approx. 110 USD; € 40 each for the adolescent and the parent). The 10-year follow-up was completed in March 2014. All procedures were approved by the Human Subjects Protection Board of the German Association of Psychology.

ZFI-sample. Out of the original N = 280 ZFI-families, n = 249 participated in the FU5 assessment. Seven families declined the participation, 2 moved to a foreign country, and we were not able to determine the addresses of n = 22 families. The responder-rate was 89% in comparison to pre in 2001/2002 and 99% in comparison to the FU4 in 2006 respectively (see Figure 1).

ZFII-sample. FU5 recruitment started 1 year after ZFI. Out of the original N = 197 families, n = 112 participated in the FU5 assessment. Five families declined the participation, and we could not determine the addresses from n = 80 families. The responder-rate was 57% in comparison to pre in 2003 and 64% in comparison to the last follow-up in 2005 respectively (see Figure 2). Reasons for the lower response rate in comparison to the ZFI-sample may be the lower socioeconomic status of the

families, the higher rate of migrants, and the more frequent change of residence (Lange et al., 2014).

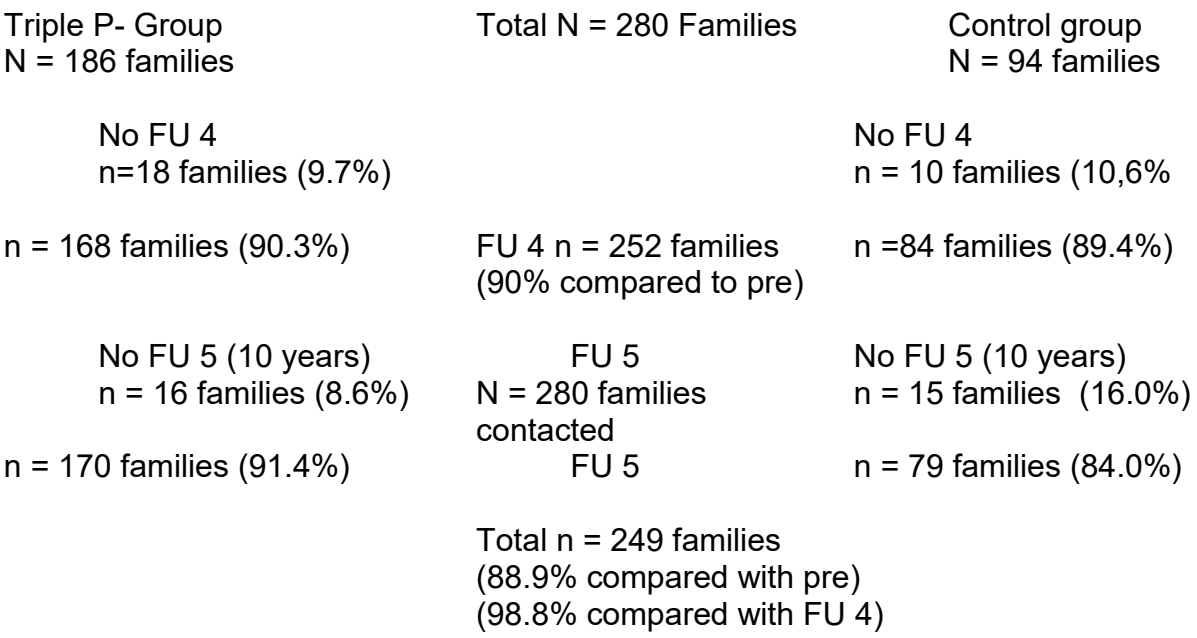

Figure 1. Study Future Family I. Flow chart of families recruitment from pre to follow-up FU 5 (10 years).

ZFII-sample. FU5 recruitment started 1 year after ZFI. Out of the original N = 197 families, n = 112 participated in the FU5 assessment. Five families declined the participation, and we could not determine the addresses from n = 80 families. The responder-rate was 57% in comparison to pre in 2003 and 64% in comparison to the last follow-up in 2005 respectively (see Figure 2). Reasons for the lower response rate in comparison to the ZFI-sample may be the lower socioeconomic status of the families, the higher rate of migrants, and the more frequent change of residence (Lange et al., 2014).

ZFIII-sample

Out of the possible 477 families at pre N = 361 participated at FU5 10 years later. In total, the response rate was 76%, with the highest rate for the T- sample (see Electronic Supplement [ESM] 1 in the appendix). Drop-outs in comparison to FU5 participants were more often single and had a lower levels of parental school grades, Kindergarden-Social-Structural Index OKS, income (all $p < .001$). In the other instruments no significant differences emerged. Families were distributed as follows: Triple P T+: N = 242; Triple P T-: N = 40, and Controlgroup CG: N = 79.

Sociodemographic Data. Mean age of mothers/fathers/adolescents (54% males) were 45.0 (SD = 4.9) / 48.1 (SD = 5.5) 14.1 (SD = 1.2) years. Fifty-two percent had 1 sibling, 34% more than 2, and 14% were the only child; 50% attended High School (13 years), and 44% a school with less than 13 years total (Haupt-, Real oder Gesamtschule). Forty-seven percent of mothers and 62% of fathers had a High School degree, and 38%/24%) a "Realschule"-degree (= 10 years of school). 19% of families were immigrants (one partner: 9%, both partners:10%). In 94% (n = 337) the biological mother were interviewed; 22% of mothers were single parents. Twelve percent (n = 41) were divorced. The family net income was equivalent to the German average.

Pre: N = 197 families with at least 1 Triple P-session

No Post n = 11 families (5.6%)

Post: n = 186 families (94.4%)

No FU 1 n = 5 families (2.7%)

FU 1: n = 181 families (91.9%)

No FU 2 n = 6 (3.3%)

FU 2: n = 175 families (88.8%)

No FU 5 n = 85 (43.1%)

FU 5: n 112 families (64.0% compared to FU 2; 56.9% compared to pre)

Figure 2. Study Future Family II. Flow chart of families' recruitment from pre to follow-up FU 5 (10 years).

Instruments

The effectiveness of Triple P was assessed using the following instruments: Parent Scale (PS, Arnold, O'Leary, Wolff & Acker, 1993; Naumann et al., 2010), Depressions-Anxiety-Stress-Scale (DASS, Lovibond & Lovibond, 1995, Köppe, 2001), Dyadic Adjustment Scale – Short Form (DAS-SF, Sharpley & Rogers, 1984, Köppe, 2001), Child Behavior Checklist - Parent report, pre-school form (CBCL 1½–5) or form for children and adolescents (CBCL 4–16, Döpfner et al., 2014), Youth Self Report YSR (Achenbach, 1991, Döpfner et al., 2014).

Scoring of CBCL 1½–5 and CBCL 4–16): At the assessments points pre, post, and 1-year follow-up the (CBCL 1½–5) was used for which unfortunately no German norms are available. From FU 2 up till FU 5 the CBCL/4–18 was used for which German norms are available (Döpfner et al., 2014). Both CBCL-forms are not directly

comparable: CBCL 1½–5 consists of 100 items, CBCL 4 – 18 of 118 items, and only 53 items are comparable. To compare the pre and FU5 data, two procedures were used: a) For the CBCL 1½–5, T-scores were calculated using the US-norms by Achenbach and Rescola (2000), for the CBCL 4–18 the German norms by Döpfner et al. (2014) were used. b) Per assessment point z-scores based on the total sample were calculated to allow longitudinal effectiveness analysis.

At Pre (CBCL 1,5–5) the scale means for the total sample of mothers (T-Range 49,4 to 51,0) corresponded - as expected – with the mean of T = 50 for a universal prevention sample. At FU5, the means were approximately 0,5 standard deviations higher (T Range = 52,8 – 55,0; see ESM 2 in the appendix). We can not clarify whether the higher scores 10 years later are due to fact that we used US-norms at pre or whether our sample experienced a slight deterioration.

Statistical analysis

Missing items (maximum 15% per scale) were imputed by the scale's mean. Effect-sizes (ES; *d*) were calculated by dividing the differences of the means by the pooled standard-deviation. Following Cohen (1988) interpretation of ES were: d> .20 small, d>.50 medium, and d> .80 large. For universal prevention studies only small effect-sizes are expected. Therefore large samples are needed to find significant effects: approximately N > 180 families for each group to achieve a d of 0.30 (Hiscock et al., 2008). Longitudinal studies with the danger of drop-out's need even more subjects to compensate the loss. In this study only the T+ group with an N = 232 achieved the target while the control-group (N = 79) and the T- group (N = 40) missed it. Because this study is – to our knowledge - the only (published) universal prevention study with a 10-year follow-up, we run descriptive analyses despite the insufficient statistical power. In terms of exploratory analyses we use effect-sizes for interpretation and refrained from using statistical tests.

Results

Prevalence rates of external and internal behavioral problems

In table ESM 3 (appendix), the prevalence rate of borderline (T = 60 - 63) and clinically relevant (T ≥ 64) children or adolescents are listed at the 10 year FU. According to the mothers' CBCL-assessment in the kindergarten age of subgroup ZF I, 17% of children were in the clinical range (adding the rates of borderline and clinical) with regard to internalizing behavior. With regard to externalizing problems

12% were in the clinical range (Heinrichs, Hahlweg et al., 2006). In the subgroup ZF II, at pre 34% of children were in the clinical range in the internalizing scale and 18% were in the abnormal range in the externalizing scale (Heinrichs, Kruse et al., 2006)

In adolescence, there was a substantial increase: In the ZF I subgroup, in the internalizing scale 28% of children were in the clinical range and 24% had abnormal score in externalizing behavior. The self-assessment of the adolescents (YSR) deviated only slightly from the mother's assessment. In the subgroup ZF II, according to the mothers CBCL-rating, the clinical rates were 31% for internalizing and 28% for externalizing behavior. The self-assessment of the adolescents deviated significantly, whereas adolescents reported external behavioral abnormalities far less frequently with a score of 17%.

Long-term effectiveness of the Triple P program

The evaluation was done in two steps: A) On the one hand, the T- families were evaluated together with the T+- families in the sense of an intention to treat evaluation, following the original randomization and as requested by the reviewers of the original papers (see Hahlweg et al., 2010; Heinrichs, Hahlweg et al., 2006, 2009, 2017). B) On the other hand, a differentiated analysis of the groups T+ (n = 232), T- (n = 40) and the control group (n = 79) was calculated.

A) *Intention to treat analysis*. The mothers' and fathers' data were compared with respect to efficacy of Triple P over 10 years. The differences in the parental difference values pre to FU5 of the intervention group (T+ and TP-) and the control group (CG) can be found in ESM 5 and ESM 6. There were only small effect sizes for both mothers and fathers (Mothers: ES < 0.23, Fathers ES < 0.15).

B) *Differential Analysis*. The intention to treat evaluations carried out so far (the ZFI sample) were very conservative as the 23% subjects who rejected Triple P were rated as participants even though they had not participated in any Triple P session. In an explorative, differential analysis, the three groups T+, T- and CG are compared below.

Table 1 shows the maternal pre and FU 5 mean values, standard deviations and the intragroup effect sizes for the two CBCL scales internal and external disorders, the Depression Anxiety Stress Scales DASS, the German adaptation of the Parenting Scale PS and the relationship questionnaire DAS-SF. The corresponding values of the fathers are listed in table 2.

Table 1
Total sample mothers: Changes from Pre to FU5 for subsamples T+, T- und CG. Means (M) and Standarddeviations (SD) and Intra-Group-Effect-sizes (Minus-ES: Deterioration).

		Pre		FU5			
Variable	N	M	SD	M	SD	Intra-Group ES	p
T+							
CBCL z-Score							
Internal.	216	0.03	1.02	0.02	0.99	0.01	
External.	218	0.08	1.01	0.02	1.04	0.07	
PS Total	216	3.24	0.55	2.94	0.59	0.54	***
DASS Total	215	25.5	17.9	21.4	17.5	0.22	***
DAS-SF	194	23.1	5.2	21.0	5.8	-0.36	***
T-							
CBCL z-Scores							
Internal.	35	-0.12	0.87	0.22	1.24	-0.31	*
External.	35	-0.29	0.93	-0.01	0.78	-0.28	*
PS Total	35	3.27	0.57	3.11	0.67	0.29	*
DASS Total	35	25.1	18.6	23.4	18.3	0.09	
DAS-SF	25	25.3	4.9	21.1	5.1	-0.71	***
CG							
CBCL z-Scores							
Internal.	71	-0.32	0.74	-0.16	0.89	-0.15	
External.	71	-0.24	0.94	-0.03	0.98	-0.21	*
PS Total	68	3.27	0.60	3.00	0.59	0.48	***
DASSTotal	70	21.8	13.0	18.7	14.8	0.17	*
DAS-SF	55	22.8	4.2	19.2	5.9	-0.61	***

Note. Internal. = CBCL Scale Internalizing, External. = CBCL Scale Externalizing; PS = Parenting Scale; DASS = Depression-Anxiety-Stress-Scale; DAS-SF = Dyadic Adjustment Scale – short form. Intragroup-Effectsize: (Pre – FU5) / SD(Difference). T-Tests for dependent samples. *p<.05, **p<.01, ***p<.001.

The descriptive assessment of the maternal intragroup effect sizes in Table 1 shows that T- and CG worsened slightly with respect to the CBCL scales (d = -0.15 to -0.31), while the T+ group remained virtually unchanged from pre to FU5 (d = 0.01, 0.07). In terms of dysfunctional parenting (PS) and psychological burden (DASS), all groups improved slightly, but most likely the T+ group. In terms of partnership satisfaction (DAS-SF), all groups worsened moderately (T-: d = -0.71, CG: d = -0.61), the T+ group, however, to a distinctly lesser extent (d = -0.36).

The results for the effect sizes of the fathers (Table 2): with regard to the CBCL scales, the group T- worsened slightly, in the groups CG and T+ no significant changes from pre to FU5 were identified. Regarding dysfunctional parenting behavior (EFB), minor improvements were measured in all three groups. The psychological burden (DASS) improved in the T+ group, but especially in the CG, the group T- worsened slightly. In terms of partnership satisfaction (FBZ-K), the T- fathers

worsened, while the T+ fathers slightly improved and CG fathers remained unchanged.

Table 2
Total sample fathers: Changes from Pre to FU5 for subsamples T+, T- und CG. Means (M) and Standarddeviations (SD) and Intra-Group-Effect-sizes (Minus-ES: Deterioration).

		Pre		**FU5**			
Variable	N	M	SD	M	SD	Intra-Gruppen ES	p
T+							
CBCL z-Scores							
Internal.	159	-0.10	0.79	0.01	0.97	-0.11	
External.	161	0.00	0.98	0.02	1.04	-0.02	
PS Total	160	3.16	0.50	2.98	0.60	0.34	***
DASS Total	161	19.5	14.6	17.1	13.8	0.16	*
DAS-SF	153	23.6	5.2	24.6	5.6	0.18	*
T-							
CBCL z-Werte							
Internal.	18	-0.13	0.75	0.04	1.07	-0.17	
External.	18	-0.37	0.87	-0.06	0.87	-0.32	
PS Total	17	3.14	0.46	2.98	0.50	0.32	
DASS Total	17	12.4	7.4	14.9	16.0	-0.17	
DAS-SF	16	24.7	4.9	23.5	6.00	-0.25	
CG							
CBCL z-Werte							
Internal.	47	-0.23	0.73	-0.18	0.81	0.05	
External.	47	-0.16	0.84	-0.18	0.72	-0.02	
PS Total	48	3.22	0.40	3.02	0.60	0.37	**
DASS Total	47	19.9	12.6	15.9	14.1	0.28	*
DAS-SF	44	23.1	5.3	23.1	6.2	0.01	

Note. Internal. = CBCL Scale Internalizing, External. = CBCL Scale Externalizing; PS = Parenting Scale; DASS = Depression-Anxiety-Stress-Scale; DAS-SF = Dyadic Adjustment Scale – short form. Intragroup-Effect-size: (Pre – FU5) / SD(Difference). T-Tests for dependent samples. *p<.05, **p<.01, ***p<.001.

Table 3 lists the maternal *Intergroup Effect Sizes* (*InterES*; difference in intragroup ES from Table 1). Compared to the control group CG, T+ showed low InterES for internalizing (0.16), externalizing disorders (0.27) and relationship satisfaction (0.25). Averaged across all variables, T+ mothers showed a mean InterES of 0.15 in comparison to the CG. Compared to the T- group, except for the psychological distress of mothers (DASS = 0.13), low InterES were observed (CBCL-I = 0.32; CBCL-E = 0.21; PS = 0.25, DAS-SF = 0.35). On average, an InterES of 0.25 was yielded, i.e. T- worsened to in a small extend compared to the T+ group. When T- was compared with the control group, very low InterES were found, while the T-group had worse results with an average of 0.12 standard deviations.

Table 3
Total sample mothers: Inter-group comparisons (d) for intra-group effect-sizes for subsamples T+, T- and CG. (Minus d: Deterioration).

Variable	T+ vs. CG			T+ vs. T-			T- vs. CG		
	d	95% KI	p	D	95% KI	p	d	95% KI	p
CBCL z-Werte									
Internal.	0.16	[-.08; .39]		0.32	[-.05; .70]	*	-0.16	[-.52; .20]	
External.	0.27	[.00; .55]	*	0.21	[-.03; .71]	*	-0.07	[-.41; .28]	
PS Total	0.06	[-.22; .33]		0.25	[-.10; .60]		-0.19	[-.63; .24]	
DASS Total	0.05	[-.24; .35]		0.13	[-.24; .51]		-0.08	[-.50; .34]	
DAS-SF	0.25	[-.55; .05]		0.35	[-.76; .06]	*	-0.10	[-.38; .58]	
Mean	0.15			0.25			-0.12		

Note. Internal. = CBCL Scale Internalizing, External. = CBCL Scale Externalizing; PS = Parenting Scale; DASS = Depression-Anxiety-Stress-Scale; DAS-SF = Dyadic Adjustment Scale – short form. *p<.05, **p<.01.

Table 4 shows the paternal *Intergroup Effect Sizes* (*InterES*; difference in intragroup ES from Table 2). The T+ group showed only a modest improvement for relationship satisfaction compared to the CG (0.18). Averaged across all variables, for T+ fathers there was virtually no change (-0.01) from pre to FU 10 years compared to the CG. Compared to the values of group T-, fathers showed low InterES for CBCL-E = 0.30 and for relationship satisfaction (DAS-SF = 0.43). On average, an InterES of 0.23 was yielded, i.e. T- worsened with a small effect-size compared to the T+ group. When comparing T- with the control group CG, there were consistent deteriorations at a very low level for CBCL-I (-0.12) and the overall PS questionnaire (-0.05), while low InterES were observed for CBCL-E (-0.34) and relationship satisfaction (-0.26) and moderate InterES were observed for psychological burden (DASS total = -0.45). Overall, the T- group deteriorated on average by InterES = 0.24 standard deviations.

Tabelle 4
Total sample fathers: Inter-group comparisons (d) for intra-group effect-sizes for subsamples T+, T- and CG. (Minus d: Deterioration)

Variable	T+ vs. CG			T+ vs. T-			T- vs. CG		
	d	95% KI	p	d	95% KI	p	d	95% KI	p
CBCL z-scores									
Internal.	-0.06	[-.38; .25]		.06	[-.45; .57]		-.12	[-.68; .43]	
External.	-0.04	[-.30; .22]		.30	[-.22; .83]		-.34	[-.77; .08]	[a]
PS Total	-0.03	[-.35; .29]		.02	[-.50; .54]		-.05	[-.60; .50]	
DASS Total	-0.11	[-.44; .21]		.33	[-.18; .84]	[a]	-.45	[-1.02; .12]	[a]
DAS-SF	.18	[-.17; .52]		.43	[-.09; .96]	[a]	-.26	[-.79; .28]	
Mean	-.01			.23			-.24		

Note. Internal. = CBCL Scale Internalizing, External. = CBCL Scale Externalizing; PS = Parenting Scale; DASS = Depression-Anxiety-Stress-Scale; DAS-SF = Dyadic Adjustment Scale – short form. [a]p<.1.

Table 5 shows the maternal CBCL abnormality rates (borderline and clinical relevant) as difference values FU 5 minus pre. With regard to internalizing disorders,

at FU 5 abnormality rates grew by 8% (CG), 9% (T+) and 18% (T-). For externalizing disorders, the abnormality rates decreased by 4% for T+, while growth was recorded for CG (13%) and T- (15%). Based on the sum of internal and external disorders, the ranking is as follows: T+: 5%, CG: 20% and T-: 33%. For adolescents, the 10-years FU in the Youth Self Report (YSR) and the Strength and Difficulties Questionnaire SDQ total, (see ESM 8) had only very low effect sizes between the groups T+, T- and CG (Mean <0.09).

Table 5
Mothers: CBCL-abnormalities (sum of borderline (T-scores 60 – 63) und clinically relevant (T-scores>/=64) in percentages; difference between pre vs. FU 5: T + (n=238), T – (n=39) und CG (n=79).

CBCL	Internalisizing	Externalizing	Sum Increase
Sample			
T +	9	4	5
T –	18	15	33
CG	8	13	20

Discussion

In Germany as well as in all industrialized countries, about 20% of children and adolescents suffer from mental disorders and psychosocial stress requiring treatment (Belfer, 2008; Kieling et al., 2011). These disorders put a burden on those affected, their families and their social environment and often take a chronic course. All of the above underpins the high social relevance of this issue (O'Connell et al., 2009). In Germany, at least 3 million children and adolescents are affected. These long-term negative consequences entail considerable costs for society and indicate that targeted prevention measures are needed in the sense of a public health approach in order to reduce the prevalence and incidence rates in the longer term.

In particular, cognitive-behavioral programs for skills-based parenting have proven to be effective with effect sizes of d = 0.35 (Weisz & Kazdin, 2010) and are considered to be the gold standard in the field of prevention of childhood mental disorders. However, there is hardly any catamnesis showing the sustainability of the interventions. Above all, there is a lack of long-term studies in the field of universal prevention, since most of the effects were only studied over a time period of one year (Hiscock et al., 2008).

The aim of this *Future Family III study* was to investigate the results of a parent-centered universal prevention program (Triple P) 10 years after its introduction. Main criteria were the reduction of external and internal child disorders,

dysfunctional parenting behavior, individual psychological stress of parents and improvement of partnership satisfaction. Two studies that have been carried out since the beginning of 2000 formed the basis of this study: In the randomized control group study Future Family I (ZFI), the 4-year efficacy of Triple P was evaluated for the universal prevention of child behavioral disorders for families with children aged 3 to 6 years (Heinrichs et al., 2005). In the uncontrolled study Future Family II ZFII (Heinrichs, Krüger et al., 2006) with a follow-up of two years, Triple P courses were offered to families from socially disadvantaged urban areas of Braunschweig.

Recruitment rates. N = 361 families were recruited 10 years later, which accounts for a utilization rate of 76% (ZFI: 89%, ZFII: 64%). Reasons for the more difficult access to the ZFII sample in comparison to the ZFI sample are likely to be found in a lower socio-economic status of the families, an increased migration background and a more frequent change of residence. The overall utilization rate, however, is to be assessed as good on the whole. It is comparable with the responder rates of the "Erlangen-Nürnberg Development and Prevention Study" by Lösel et al. (2013) and the first KIGGS follow-up survey in which the responder rate was 73% (Hölling et al., 2014), and also comparable with regards to the rate of explicit rejection to a renewed participation (KIGGS 4%; ZFIII 3%).

Efficacy of Group Triple P

When assessing the results, it should be remembered that universal prevention applies to all families in a population regardless of parenting behavior and child behavioral disorders. Universal interventions therefore generally lead to smaller effect sizes than, for example, indicated prevention studies, in which there are already notable problems in child behavior and also in parenting behavior. These considerations show that it is difficult to demonstrate effects of universal preventive measures, particularly with regard to long-term effects.

However, for health policy objectives and actual reductions in the incidence and prevalence rates of mental disorders in children, it is essential to demonstrate that short preventive measures are not just a "flash in the pan" but can have a lasting impact on the mental health of children. As a rule of thumb, small effect sizes are found for universal prevention (0.20 - 0.49), moderate for selective prevention (0.50 - 0.80) and large for indicated prevention (> 0.80) are considered good (Heinrichs & Hahlweg, 2009).

In universal prevention studies, therefore, about 180 families per study arm are needed in order to be able to significantly detect such minor effects (Hiscock et al., 2008). In this study, the necessary group size could only be achieved in the T+ group, but not in the control group and the T- group. In addition, not all families completed the entire questionnaires. However, since our study is the only study with a universal approach in Germany, apart from the "Erlangen-Nürnberg Development and Prevention Study" by Lösel et al. (2013), an exploratory, descriptive evaluation seemed to make sense despite the lack of power.

Intention to treat analysis. As in the previous publications on ZFI (Hahlweg et al., 2010, Heinrichs, Hahlweg et al., 2006, 2009, 2017), the original randomization was maintained in the sense of an intention to treat analysis and the T+ and T- families were evaluated together. This strategy is very conservative as the 23% of parents who rejected to participate in any sessions of Group Triple P were rated as participants. As expected from the above considerations, mothers and fathers showed no significant differences and the effect sizes were very small.

Intragroup Effect Sizes. Therefore, a differential analysis strategy was chosen for the three groups T +, T-, and CG to determine the "pure" effects of the Triple P intervention. Additionally, for the first time to our knowledge, a long-term examination of those refusing parenting training could be carried out. The intragroup effect sizes of the *mothers* displayed a substantially uniform picture: the T+ Group achieved the best results – at a low level – followed by the control group and the T- group. With regard to the CBCL measures, the T- und CG groups worsened slightly, while the T+ group did not change from pre to FU5. In terms of dysfunctional parenting behavior (PS) and psychological burden (DASS), all groups improved slightly; in terms of partnership satisfaction (DAS-SF), all groups worsened moderately, but group T+ to a lesser extent.

The results for the fathers' intragroup effect sizes should be judged against the background of their very low intervention participation, as only 6% attended Group Triple P sessions. This low paternal participation rate and the few significant results have been reported since the 1-year-FU (Heinrichs, Hahlweg et al., 2006) and also in other studies (Bodenmann, Cina, Ledermann & Sanders, 2008; Sanders, Markie-Dadds, Tully & Bor, 2000). Regarding the ranking of the groups, the results of the

fathers are similar to those of the mothers: T+/CG, followed by T-, but the differences are less pronounced.

Intergroup Effect Sizes (InterES). Due to the small sample sizes (and therefore insufficient statistical power), a analysis of variance was omitted and InterES calculated (difference in intragroup ES). For adolescents, only the 10-years follow-up data were available. In contrast to parents, there were no differences between the T+, T- and CG groups in either the Youth Self Report subscales nor the Strength and Difficulties Scale (SDQ).

Comparison of T+ with CG-families. From the point of view of the T+ mothers there were small InterES (improvements) for internalizing (0.16) and externalizing disorders (0.27), and relationship satisfaction (0.25) compared to the control group CG. Averaged across all variables a mean InterES of d = 0.15 was found for T+ mothers. Thus, from the main criteria, the CBCL variables still showed low effect sizes that met the expectations for universal preventive studies, but there were no significant InterES for parenting competence. However, it must be taken into account that the PS questionnaire has been validated primarily for children aged 3-12 years. Whether it is also useful for adolescents has not yet been investigated. For the DASS, which assesses psychological burden, no significant InterES were found. However, it should be emphasized, that the relationship quality of T+ mothers improved to a minor extent – although this issue was not addressed in the Triple P courses. This may be an effect that could be attributed to an improvement in adolescent's behavior. There were no differences between T+ and CG among fathers.

Comparison of T- with T+ and CG families. Compared to the values of group T-, mild to moderate InterES were found, apart from the psychological burden of mothers and fathers. On average, there was an InterES of d = 0.25 for mothers and 0.23 for fathers, i.e. T- deteriorated slightly compared to the T+ group. When comparing T- with the control group, non-significant InterES (-0.12) were observed for mothers and low InterES (-0.24) for fathers, with the T- group scoring worse than the control group.

Overall, this means that from the point of view of mothers and fathers, those refusing Triple P show significant deteriorations after 10 years compared to those participating in Triple P. This is also shown by the categorical CBCL abnormality rates of the sum of internal and external disorders: After 10 years, T- mothers

reported an increase of 33% compared to the pre values, CG mothers an increase of 20% and T+ mothers an increase of only 6%.

T- families therefore represent a risk group that should be examined more intensively in the future. Interestingly, the T- families had the highest retention rate, namely 95% after 10 years, followed by the T+ families (90%) and the control families (84%). We would have expected a rate comparable to the control group as both groups did not receive any intervention. When taking into account the higher rate of adolescents with conspicuous CBCL values at FU5, the strong motivation of T- families to participate in the 10-years follow-up might be explained by the parents' wish to receive more professional information and/or support in view of their dependents' alarming psychological disorders.

The strengths of this study lie in the long follow-up period with good retention rates, the investigation from three perspectives: mother, father and adolescents and the first examination of individuals who rejected the intervention over a long period of 10 years.

The results are subject to some important limitations. Above all, the insufficient sample size in the T- and control group did not allow a meaningful significance test. However, this would have necessitated recruiting an additional 100 families for the control group at pre-time – and for the T- group with a rejection rate of 25%, a total of 600 additional families would have been needed in order to achieve sufficient statistical power.

The long-term study required a change in the age-limited versions of central measuring instruments and, as a result, a change in the standardization samples, in some cases relying on standardization samples outside of Germany. It is therefore not possible to determine to what extent the observed trends can also be attributed to these methodological causes. These potential shortcomings do not take effect, however, when comparing the different subgroups.

Despite the encouraging results on the effectiveness of interventions to increase parenting competencies, further research is necessary. Lösel and Runkel (2013) pointed out the complexity of the research area and listed the potential factors influencing the effects of prevention programs in a model: In addition to participant characteristics (such as risk level, age, motivation), context variables (relationship quality, organizational characteristics), program characteristics (program content, intensity / dosage), evaluation methodology plays an important role (sample size,

practice evaluation, length of follow-up). This model is suitable as a concept for future research. From their point of view "...there is no urgent need for constantly developing new prevention programs. Rather, the empirically founded interventions should be systematically implemented and further developed. More controlled long-term evaluations of effectiveness and process evaluations are required in order to better reach the target groups and ensure quality of implementation. In both fields, it is increasingly necessary to move from small model projects to large-scale, multi-center practice evaluations" (Lösel & Runkel, 2013, S. 234).

Increased research into the dissemination of evidence-based prevention programs into routine care is also important in order to provide as many families as possible with effective services. From the study by Frantz and Heinrichs (2015) important recommendations for action can be derived that facilitate a successful dissemination of evidence-based programs into practice. These are summarized in a checklist, which can be found in the above mentioned publication, and which provides valuable information for developers and users.

In principle, it would be desirable if such evidence-based interventions were available to as many families as possible by now. Unfortunately, family-oriented preventive measures are currently not covered by the Prevention Act ("Präventionsgesetz" - PrävG) adopted on 10 July 2015 so that families have to cover the costs themselves. This certainly represents a barrier to participation in appropriate parenting courses, as it mainly affects young families who are financially burdened by children. A change in the law that would allow at least partial funding seems urgently necessary, since early intervention promises the best cost-benefit ratio (Cierpka, Stasch & Groß, 2007; Heckman, 2008; Lösel & Runkel, 2013). Hopefully this study on the encouraging long-term effects will contribute to launch such initiatives.

The results of this 10-year follow-up confirm the findings of the efficacy studies to date. While so far only follow-up periods of up to three years have been investigated, with this explorative study results are available for the first time after 10 years. This study thus supports the recommendation derived from the findings to date (e.g. United Nations Office on Drugs and Crime, 2009; World Health Organization, 2009) to offer parenting support as early as possible.

References

Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington: Department of Psychiatry, University of Vermont.

Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for the ASEBA Preschool Form & Profiles*. Burlington: University of Vermont, Department of Psychiatry.

Arbeitsgruppe Deutsche Child Behavior Checklist (2000). *Elternfragebogen für Klein- und Vorschulkinder (CBCL 1 ½ -5)*. Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik (KJFD).

Arnold, D. S., O'Leary, S. G., Wolff, L. S., & Acker, M. M. (1993). The Parenting Scale: A measure of dysfunctional parenting in discipline situations. *Psychological Assessment, 5*, 131-144.

Bäse, B. (1995). *Die sozial-räumliche Gliederung der Stadt Braunschweig. Methodik und Durchführung sozial-geographischer Analyse im städtischen Wohnumfeld auf der Grundlage des Zensus 1987.* Braunschweig: Technische Universität Braunschweig.

Belfer, M. L. (2008). Child and adolescent mental disorders: The magnitude of the problem across the globe. *Journal of Child Psychology and Psychiatry, 49*(3), 226-236.

Bodenmann G, Cina A, Ledermann T, Sanders MR: The efficacy of the Triple P-Positive Parenting Program in improving parenting and child behavior: A comparison with two other treatment conditions**.** *Behaviour Research and Therapy* 2008, 46:411-427.

Caminis, A., Henrich, C., Ruchkin, V., Schwab-Stone, M., & Martin, A. (2007). Psychosocial predictors of sexual initiation and high-risk sexual behaviors in early adolescence. *Child and Adolescent Psychiatry and Mental Health, 1*(14), 14.

Cierpka, M., Stasch, M. & Groß, S. (2007). *Expertise zum Stand der Prävention/Frühprävention in der frühen Kindheit in Deutschland*. Köln: Bundeszentrale für Geundheitlich Aufklärung BZGA.

Destatis, 2014; www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Gesundheit/ Gesundheitsausgaben/Gesundheitsausgaben.html, 16.01.2017

Döpfner, M., Plück, J., Kinnen, C., Arbeitsgruppe Deutsche Child Behavior Checklist (2014). *CBCL/6-18R, TRF/6-18R, YSR/11-18R. Deutsche Schulalter-Formen der Child Behavior Checklist von Thomas M. Achenbach. Elternfragebogen über das Verhalten von Kindern und Jugendlichen (CBCL/6-18R), Lehrerfragebogen über das Verhalten von Kindern und Jugendlichen (TRF/6-18R), Fragebogen für Jugendliche (YSR/11-18R)*. Göttingen: Hogrefe.

Eisner, M., Nagin, D., Ribeaud, D., & Malti, T. (2012). Effects of a universal parenting program for highly adherent parents: A propensity score matching approach. *Prevention Science, 13*(3), 252-266.

Englund, M. M., Egeland, B., Oliva, E. M., & Collins, W. A. (2008). Childhood and adolescent predictors of heavy drinking and alcohol use disorders in early adulthood: A longitudinal developmental analysis. *Addiction, 103*, 23-35.

Farrington, D. P., & Welsh, B. C. (2007). *Saving children from a life of crime: Early risk factors and effective interventions.* Oxford: Oxford University Press.

Fernandez, M., & Eyberg, S. (2009). Predicting treatment and follow-up attrition in Parent–Child Interaction Therapy. *Journal of Abnormal Child Psychology, 37*, 431–441.

Forgatch, M. S., & Patterson, G. R. (2010). Parent Management Training — Oregon Model: An intervention for antisocial behavior in children and adolescents. In J. R.Weisz, & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 159–178) (2nd ed.). New York, NY: Guilford.

Frantz, I. & Heinrichs, N. (2015). Implementation von in der Forschung untersuchten Präventionsprogrammen in die Praxis. Akzeptanz und Barrieren. *Zeitschrift für Klinische Psychologie und Psychotherapie, 44*(1), 56-61.

Furlong, M., McGilloway, S. M., Bywater, T., Hutchings, J., Smith, S. M. & Donnelly, M. (2012). Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years. *Cochrane Database Systematic Review, 15*, doi: 10.1002/14651858.CD008225.pub2.

Gustavsson, A., Svensson, M., Jacobi, F., Allgulander, C., Alonso, J., Beghi, E. et al. (2011). Cost of disorders of the brain in Europe 2010. *European Neuropsychopharmacology, 21*(10), 718-779.

Hahlweg, K. (2013). Langzeitwirksamkeit und Dissemination von Elternkompetenztrainings.. In L. Corell & J. Lepperhoff (Hrsg.), *Frühe Bildung in der Familie. Perspektiven der Familienbildung*. (S.239-252).Weinheim, Basel: Beltz Juventa.

Hahlweg, K. & Heinrichs, N. (2007). Prävention psychischer Störungen bei Kindern und Jugendlichen: Evidenzbasierte Elterntrainings in Deutschland. In W. Kirch, B. Badura & H. Pfaff (Hrsg.), *Prävention und Versorgungsforschung*. (S. 669-690). Heidelberg: Springer.

Hahlweg, K., Heinrichs, N., Naumann, S., Kuschel, A. & Bertram, H. (2010). Two-year outcome of universal prevention through a positive parenting program: Is it worth the effort? *Child and Adolescent Psychiatry and Mental Health, 4,* 14-27.

Heckman, J. J. (2008): Schools, skills and synapses. *Economic Inquiry*, *46*, 289–324.

Heinrichs, N. & Hahlweg, K. (2009). Elterntrainings zur Steigerung der Erziehungskompetenz. In S. Schneider & J. Margraf (Hrsg.), *Lehrbuch der Verhaltenstherapie, Band 3* (S. 255-276). Berlin, Heidelberg: Springer.

Heinrichs, N., Bertram, H., Kuschel, A. & Hahlweg, K. (2005). Parent recruitment and retention in a universal prevention program for child behavior and emotional problems: Barriers to research and program participation. *Prevention Science*, *6*, 275-286

Heinrichs, N., Hahlweg, K., Bertram, H., Kuschel, A., Naumann, S. & Harstick, S. (2006). Die langfristige Wirksamkeit eines Elterntrainings zur universellen Prävention kindlicher Verhaltensstörungen. Ergebnisse aus Sicht der Mütter und Väter. *Zeitschrift für Klinische Psychologie und Psychotherapie, 35*, 82-95.

Heinrichs, N., Hahlweg, K., Naumann, S., Kuschel, A., Bertram, H., & Ständer, D. (2009). Universelle Prävention kindlicher Verhaltensstörungen mithilfe einer elternzentrierten Maßnahme. *Zeitschrift für Klinische Psychologie und Psychotherapie, 38*(2), 79-88.

Heinrichs, N. & Jensen-Doss, A. (2010). The effects of incentives on families' long-term outcome in a parenting program. *Journal of Clinical Child & Adolescent Psychology, 39*: 5, 705 — 712; DOI: 10.1080/15374416.2010.501290.

Heinrichs, N., Krüger, S. & Guse, U. (2006). Der Einfluss von Anreizen auf die Rekrutierung von Eltern und auf die Effektivität eines Elterntrainings. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 35*, 97-108.

Henry-Huthmacher, C. (2008). Eltern unter Druck. Die wichtigsten Ergebnisse der Studie. In C. Henry-Huthmacher & M. Borchard (Hrsg.), *Eltern unter Druck. Selbstverständnisse, Befindlichkeiten und Bedürfnisse von Eltern in verschiedenen Lebenswelten.* Stuttgart: Lucius & Lucius.

Hiscock, H., Bayer, J. K., Price, A., Ukoumunne, O. C., Rogers, S., & Wake, M. (2008). Universal parenting programme to prevent early childhood behavioural problems: Cluster randomised trial. *British Medical Journal 336*, 318-321.

Hölling, H., Schlack, R., Petermann, F., Ravens-Sieberer, U., & Mauz, E. (2014). Psychische Auffälligkeiten und psychosoziale Beeinträchtigungen bei Kindern und Jugendlichen im Alter von 3 bis 17 Jahren in Deutschland–Prävalenz und zeitliche Trends zu 2 Erhebungszeitpunkten (2003-2006 und 2009-2012). *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, 57*(7), 807-819.

Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology, 36*(4), 567-589.

Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., et al. (2011). Child and adolescent mental health worldwide: Evidence for action. *The Lancet, 378*(9801), 1515-1525.

Klasen, H., Woerner, W., Rothenberger, A. & Goodman, R. (2003). Die deutsche Fassung des Strengths and Difficulties Questionnaire (SDQ-Deu)–Übersicht und Bewertung erster Validierungs- und Normierungsbefunde. *Praxis der Kinderpsychologie und Kinderpsychiatrie, 52,* 491-502.

Köppe, E. (2001). *Glückliche Eltern, liebe Kinder: Auswirkungen von Partnerschaft und psychischer Symptomatik der Eltern auf das Verhalten ihrer Kinder*. Unveröffentlichte Dissertation, Technische Universität Braunschweig.

Kuschel, A., Lübke, A., Köppe, E., Miller, Y., Hahlweg, K. & Sanders, M. R. (2004). Häufigkeit psychischer Auffälligkeiten und Begleitsymptome bei drei- bis sechsjährigen Kindern: Ergebnisse der Braunschweiger Kindergartenstudie. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 32*, 97-106.

Lange, M., Butschalowsky, H. G., Jentsch, F., Kuhnert, R., Rosario, A. S., Schlaud, M., & Kamtsiuris, P. (2014). Die erste KiGGS-Folgebefragung (KiGGS Welle 1). *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, *57*(7), 747-761.

Lösel, F. & Runkel, D. (2013). Prävention von kindlichen Verhaltensproblemen in Familien. Ergebnisse und Schwierigkeiten der Evaluation. In L. Corell & J. Lepperhoff (Hrsg.), *Frühe Bildung in der Familie. Perspektiven der Familienbildung*. (S.224-238).Weinheim, Basel: Beltz Juventa.

Lösel, F., Stemmler, M., & Bender, D. (2013). Long-term evaluation of a bimodal universal prevention program: Effects on antisocial development from kindergarten to adolescence. *Journal of Experimental Criminology, 9*(4), 429-449.

Lundahl, B., Risser, H. J., & Lovejoy, M. C. (2006). A meta-analysis of parent training: Moderators and follow-up effects. *Clinical Psychology Review, 26*, 86-104.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety inventories. *Behaviour Research and Therapy, 33*, 335-343.

Naumann, S., Bertram, H., Kuschel, A., Heinrichs, N., Hahlweg, K. & Döpfner, M. (2010). Der Erziehungsfragebogen (EFB). Ein Fragebogen zur Erfassung elterlicher Verhaltenstendenzen in schwierigen Erziehungssituationen. *Diagnostica, 56*, 144-157.

Nowak, C., & Heinrichs, N. (2008). A comprehensive meta-analysis of Triple P-Positive Parenting Program using hierarchical linear modeling: Effectiveness and moderating variables. *Clinical Child and Family Psychology Review, 11*(3), 114-144.

O'Connell, M. E., Boat, T., & Warner, K. E. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington, DC: National Academies Press.

Pingault, J., Côté, S., Galéra, C., Genolini, C., Falissard, B., Vitaro, F., & Tremblay, R. (2013). Childhood trajectories of inattention, hyperactivity and oppositional behaviors and prediction of substance abuse/dependence: A 15-year longitudinal population-based study. *Molecular Psychiatry, 18*(7), 806-812.

Piquero, A. R., Jennings, W. G., Diamond, B., Farrington, D. P., Tremblay, R. E., Welsh, B. C., & Reingle Gonzalez, J. M. (2016). A meta analysis update on the effects of early family/parent training programs on antisocial behavior and delinquency. *Journal Experimental Crimonology, 12*, 229-248.

Propp, O., Schilder, A., Hahlweg, K., Hannighofer, J. & Schulz, W. (2014). Übereinstimmung von Mutter-Kind-Aussagen und deren Einflussfaktoren am Beispiel des Kinder-DIPS in

der Diagnostik psychischer Störungen im Jugendalter. *Zeitschrift für Klinische Psychologie und Psychotherapie, 43*(2), 92-103.

Ravens-Sieberer, U., Wille, N., Bettge, S. & Erhart, M. (2007). Psychische Gesundheit von Kindern und Jugendlichen in Deutschland: Ergebnisse aus der BELLA-Studie im Kinder und Jugendgesundheitssurvey (KiGGS). *Bundesgesundheitsblatt–Gesundheitsforschung–Gesundheitsschutz, 50*, 871-878.

Rustenbach, S.J. (2003). *Metaanalyse. Eine anwendungsorientierte Einführung.* Bern: Huber.

Sanders, M. R. (2012). Development, evaluation, and multinational dissemination of the Triple P-Positive Parenting Program. *Annual Review of Clinical Psychology, 8*, 1-35.

Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014). The Triple P-Positive Parenting Program: A systematic review and meta-analysis of a multi-level system of parenting support. *Clinical Psychology Review, 34*(4), 337-357.

Sanders, M. R., Markie-Dadds, C., Tully, L. A., Bor, W. (2000). The Triple P-Positive Parenting Program: A comparison of enhanced, standard, and selfdirected behavioural family intervention for parents of children with early onset conduct problems**.** *Journal of Consulting and Clinical Psychology,* 68, 624-640.

Sandler, I., Schoenfelder, E., Wolchik, S., & MacKinnon, D. (2011). Long-term impact of prevention programs to promote effective parenting: Lasting effects but uncertain processes. *Annual Review of Psychology, 62*, 299-330.

Schneider, S., Unnewehr, S. & Margraf, J. (2009). *Kinder-DIPS: Diagnostisches Interview bei psychischen Störungen im Kindes- und Jugendalter.* Heidelberg: Springer.

Sharpley, C. F., & Rogers, H. J. (1984). Preliminary validation of the Abbreviated Spanier Dyadic Adjustment Scale: Some psychometric data regarding a screening test of marital adjustment. *Educational and Psychological Measurement, 44*, 1045-1049.

Stattin, H., Enebrink, P., Özdemir, M., & Giannotta, F. (2015). A national evaluation of parenting programs in Sweden: The short-term effects using an RCT effectiveness design. *Journal of Consulting and Clinical Psychology, 83*(6), 1069-1084.

United Nations Office on Drugs and Crime [UNDOC] (2009). *Guide to implementing family skills training programmes for drug abuse prevention*. Vienna, Austria: United Nations.

Webster-Stratton, C. (1998). Preventing conduct problems in Head Start children: Strengthening parenting competencies. *Journal of Consulting and Clinical Psychology, 66*, 715–730.

Weiss, M., Schmucker, M., & Lösel, F. (2015). Meta-Analyse zur Wirkung familienbezogener Präventionsmaßnahmen in Deutschland. *Zeitschrift für Klinische Psychologie und Psychotherapie, 44*, 27-44.

Weisz, J. R., & Kazdin, A. E. (2010). *Evidence-based psychotherapies for children and adolescents*. New York, NY: Guilford Press.

Wilson, P., Rush, R., Hussey, S., Puckering, C., Sim, F., Allely, C. S.,et al. (2012). How evidence-based is an 'evidence-based parenting program'? A PRISMA systematic review and meta-analysis of Triple P. *BMC Medicine, 10*(1), 130.

WHO World Health Organisation (2009). *Preventing violence through the development of safe, stable and nurturing relationships between children and their parents and caregivers. Series of briefings on violence prevention: The evidence*. Geneva, Switzerland: World Health Organisation.

Wymbs, B. T., Pelham Jr., W. E., Molina, B. S. G., Gnagy, E. M., Wilson, T. K., & Greenhouse, J. B. (2008). Rate and predictors of divorce among parents of youth with ADHD. *Journal of Consulting and Clinical Psychology, 76*, 735 – 744.

ESM Table 1

N of participants at FU5 (10 year follow-up). For the total sample (ZF I + ZF II) and subsamples ZF I (T+, T-, KG) and ZF II

	Participated at FU5				
	yes		no		
Sample	n	%	n	%	n
ZF I T+	130	90,3	14	9,7	144
ZF I T-	40	95,2	2	4,8	42
ZF I CG	79	84,0	15	94	94
ZF I Total	249	89,8	31	10,2	280
ZF II Total	112	56,9	85	43,1	197
Total	**361**	**75,7**	**116**	**24,3**	**477**

Note. TP+ = Triple P offer accepted, TP- = Triple P-offer declined, CG = Controlgroup

ESM Table 2

Mothers, only FU5 participants: CBCL T-scores at pre (CBCL 1,5-5) and at FU5 (CBCL 4-18) für Total sample, Controlgroup (CG), Triple P accepted (T+) und Triple P-declined (T-). Sample in brackets.

		Total (N=326)		CG (n=71)		T+ (n=219)		T- (n=39)	
CBCL T-Scores	Time	M	SD	M	SD	M	SD	M	SD
Total	Prä	49,9	10,2	47,9	8,7	50,7	10,5	48,6	11,4
	FU 5	53,4	10,0	52,3	9,9	53,7	10,0	54,2	10,6
Internal	Prä	51,0	10,3	48,5	9,0	51,9	10,6	50,9	11,0
	FU5	55,0	9,7	53,6	9,0	55,2	9,7	56,4	10,9
External	Prä	49,4	9,8	47,6	9,1	50,5	9,9	46,4	9,8
	FU 5	52,8	10,3	52,3	10,7	52,9	10,4	53,2	9,5

Note. M = Mean, SD = Standarddeviation

ESM Table 3

Comparison of mother (CBCL 4-18) and youth selfreport (YSR) for subsamples of ZF I (n = 249) and ZF II (n = 112) and the total sample ZF III (N=361) at FU5

Behavioral disorders	Mother ZF I % (n)	Youth ZF I % (n)	Mother ZFII % (n)	Youth ZF II % (n)	Mothe ZF III % (n)	Youth ZF III % (n)
internalizing clinical range[1]	16 (36)	17 (41)	21 (19)	11 (12)	17 (55)	15 (53)
internalizing borderline[2]	12 (27)	9 (23)	10 (9)	14 (16)	11 (22)	11 (39)
externalizing clinical range[3]	12 (28)	6 (14)	15 (14)	8 (9)	13 (42)	6 (23)
Externalizing borderline[4]	12 (27)	13 (31)	13 (12)	6 (7)	12 (39)	11 (38)

Note. CBCL = Child Behavior Checklist, YSR = Youth Self Report. [1]Internalizing clinical range: CBCL- scores >13 for girls, >12 for boys. YSR-scores >20 for girls, > 16 for boys. [2]Internalizing borderline: CBCL 10 to 12 for girls, 8 to 11 for boys. YSR: 15 to 19 for girls, 12 to 15 for boys. [3]Externalizing clinical range: CBCL >14 for girls, >18 for boys. YSR: >20 for girls, >22 for boys. [4] Externalizing borderline: CBCL 10 to 13 for girls, 13 to 17 for boys. YSR: 16 to 19 for girls, 17 to 21 for boys.

ESM Tabelle 4

Total sample fathers: Inter-group comparisons (d) for intra-group effect-sizes for subsamples T+, T- and CG. (Minus-d: Deterioration).

Variable	**T+ vs. CG**			**T+ vs. T-**			**T- vs. CG**		
	d	95% KI	p	d	95% KI	p	d	95% KI	p
CBCL z-scores									
Internal.	-0.06	[-.38; .25]		.06	[-.45; .57]		-.12	[-.68; .43]	
External.	-0.04	[-.30; .22]		.30	[-.22; .83]		-.34	[-.77; .08]	[a]
EFB Gesamt	-0.03	[-.35; .29]		.02	[-.50; .54]		-.05	[-.60; .50]	
DASS Gesamt	-0.11	[-.44; .21]		.33	[-.18; .84]	[a]	-.45	[-1.02; .12]	[a]
FBZ-K	.18	[-.17; .52]		.43	[-.09; .96]	[a]	-.26	[-.79; .28]	
Mean	-.01			.23			-.24		

Note. Internal. = CBCL Scale Internalizing, External. = CBCL Scale Externalizing; PS = Parenting Scale; DASS = Depression-Anxiety-Stress-Scale; DAS-SF = Dyadic Adjustment Scale – short form. [a]P>.10.

ESM Table 5
Mothers: Comparison of the Intragroup-Effectsizes fro, pre to FU5 between Triple P and controlgroup CG

	Triple P			CG			
Variable	N	M	SD	N	M	SD	d
CBCL 4-18 - Z-scores							
Internalisierend	251	-.04	1.05	71	-.15	.80	-.11
Externalisierend	253	.02	1.03	71	-.21	.87	-.23
DASS - Total	250	.20	1.04	70	.17	.83	-.03
PS - Total	251	.50	.98	68	.48	1.08	-.02
DAS-SF	219	.40	.99	55	.61	1.05	.21

Note. N = sample, M = Mean of Intragroup-Effectsize d, SD = Standarddeviation of Intra-droup-Effectsize, d = g_{Hedges}.

ESM Table 6
Fathers: Comparison of the Intragroup-Effectsizes for pre to FU5 between Triple P and controlgroup CG

	Triple P			CG			
Variable	N	M	SD	N	M	SD	d
CBCL 4-18 - Z-scores							
Internalisierend	177	-.12	1.04	47	-.05	.85	.07
Externalisierend	179	-.05	1.07	47	.02	.69	.07
DASS - Total	178	.13	1.02	47	.28	.94	.15
PS - Total	177	.34	1.03	48	.37	.89	.03
DAS-SF	169	-.14	1.02	44	-.01	.94	.14

Note. N = sample, M = Mean of Intragroup-Effectsize d, SD = Standarddeviation of Intra-droup-Effectsize, d = g_{Hedges}.

ESM Table 7
Total sample adolescents: Means (M), Standarddeviations (SD), and Inter-Group-comparisons (d) (T+ vs. CG, T+ vs. T- und T- vs. CG).

Variable	T+			T-			KG		
	N	M	SD	N	M	SD	N	M	SD
CBCL-YSR T-scores									
Internalisierend	241	54.4	8.7	40	55.5	10.1	77	53.7	9.3
Externalisierend	241	51.6	7.9	40	52.2	6.8	77	53.3	8.1
SDQ – Total	241	9.5	4.52	40	9.8	4.37	77	9.5	5.0

Variable	**T+ vs. CG**			**T+ vs. T-**			**T- vs. CG**		
	d	95% KI	p	d	95% KI	p	d	95% KI	p
CBCL-YSR T-Werte									
Internalisierend	-0.08	[-.34; .18]		0.12	[-.22; .45]		-0.18	[-.56; .20]	
Externalisierend	0.21	[-.05; .47]	*	0.08	[-.26; .41]		0.14	[-.24; .52]	
SDQ – Gesamt	0.01	[-.26; .25]		0.08	[-.26; .41]		-0.08	[-.46; .31]	
Mittelwert	0.05			.09			-.04		

Note: SDQ = Strength and Difficulties Questionnaire. [a]p<.1, *p<.05, **p<.01, ***p<.001.